


RESEARCH ARTICLE

Open Access



The plasmid-mediated evolution of the mycobacterial ESX (Type VII) secretion systems

Mae Newton-Foot^{*} , Robin Mark Warren, Samantha Leigh Sampson, Paul David van Helden and Nicolaas Claudius Gey van Pittius

Abstract

Background: The genome of *Mycobacterium tuberculosis* contains five copies of the ESX gene cluster, each encoding a dedicated protein secretion system. These ESX secretion systems have been defined as a novel Type VII secretion machinery, responsible for the secretion of proteins across the characteristic outer mycomembrane of the mycobacteria. Some of these secretion systems are involved in virulence and survival in *M. tuberculosis*; however they are also present in other non-pathogenic mycobacteria, and have been identified in some non-mycobacterial actinomycetes. Three components of the ESX gene cluster have also been found clustered in some gram positive monoderm organisms and are predicted to have preceded the ESX gene cluster.

Results: This study used *in silico* and phylogenetic analyses to describe the evolution of the ESX gene cluster from the WXG-FtsK cluster of monoderm bacteria to the five ESX clusters present in *M. tuberculosis* and other slow-growing mycobacteria. The ancestral gene cluster, ESX-4, was identified in several nonmycomembrane producing actinobacteria as well as the mycomembrane-containing *Corynebacteriales* in which the ESX cluster began to evolve and diversify. A novel ESX gene cluster, ESX-4_{EVOL}, was identified in some non-mycobacterial actinomycetes and *M. abscessus* subsp. *bolletii*. ESX-4_{EVOL} contains all of the conserved components of the ESX gene cluster and appears to be a precursor of the mycobacterial ESX duplications. Between two and seven ESX gene clusters were identified in each mycobacterial species, with ESX-2 and ESX-5 specifically associated with the slow growers. The order of ESX duplication in the mycobacteria is redefined as ESX-4, ESX-3, ESX-1 and then ESX-2 and ESX-5. Plasmid-encoded precursor ESX gene clusters were identified for each of the genomic ESX-3, -1, -2 and -5 gene clusters, suggesting a novel plasmid-mediated mechanism of ESX duplication and evolution.

Conclusions: The influence of the various ESX gene clusters on vital biological and virulence-related functions has clearly influenced the diversification and success of the various mycobacterial species, and their evolution from the non-pathogenic fast-growing saprophytic to the slow-growing pathogenic organisms.

Keywords: ESX, ESAT-6, Evolution, Mycobacterium, Plasmid, Type VII secretion system

Background

The genome of *Mycobacterium tuberculosis* contains five ESX (or ESAT-6) gene clusters, named ESX-1, -2, -3, -4 and -5, which encode the Esx and PE/PPE proteins, various ATPases, membrane proteins, the mycosin proteases and other ESX-associated proteins [1, 2]. The ESX gene

clusters have been the topic of extensive research following the discovery that the primary attenuating deletion of *M. bovis* BCG, region of difference 1 (RD1), includes part of ESX-1 [3–5]. The proteins encoded in each of the ESX gene clusters have been predicted to form dedicated protein secretion systems, the ESX secretion systems, which have since been defined as a Type VII secretion machinery responsible for the secretion of, amongst others, the Esx, PE and PPE proteins encoded in them, across the outer mycomembrane [6, 7].

* Correspondence: maen@sun.ac.za

Division of Molecular Biology and Human Genetics, Department of Biomedical Sciences, DST/NRF Centre of Excellence for Biomedical Tuberculosis Research/SAMRC Centre for Molecular and Cellular Biology, Faculty of Medicine and Health Sciences, Stellenbosch University, Stellenbosch, South Africa



The functions of the five *M. tuberculosis* ESX secretion systems appear to be distinct. ESX-1 is associated with virulence in *M. tuberculosis* [8–10], where it is involved in the inhibition of T-cell responses and phagosome maturation [11, 12], and assists in the escape of mycobacteria from the macrophage vacuole by ESAT-6-mediated perforation of the vacuolar membrane [13–16]. ESX-5 has also been linked to *M. tuberculosis* pathogenicity and is involved in modulating the host immune responses to maintain a persistent infection [15, 17, 18]. ESX-5 has furthermore been linked to the uptake of nutrients by increasing outer-membrane permeability in the slow-growing mycobacteria [19]. ESX-3 is essential for the *in vitro* growth of *M. tuberculosis* [20, 21], and is involved in divalent cation (iron and zinc) homeostasis [22, 23], and specifically iron uptake via the mycobactin iron acquisition pathway [21, 24]. The functions of ESX-2 and ESX-4 remain unknown.

The ESX gene clusters occur throughout the genus *Mycobacterium*. A previous study has proposed the order of duplication of the ESX gene clusters to be ESX-4, -1, -3, -2 and then -5, with ESX-5 exclusively associated with the slow-growing mycobacteria [2]. The non-pathogenic, fast-growing mycobacterium, *M. smegmatis*, contains three of the five *M. tuberculosis* ESX gene clusters, ESX-1, -3 and -4 [2]. In *M. smegmatis*, ESX-1 has been shown to be involved in conjugal DNA transfer [25, 26]. ESX-3 is also involved in iron homeostasis, however it has not been directly linked to zinc homeostasis, and is not essential in this organism [27]. Although there are distinct contrasts in the functions of these secretion systems in *M. smegmatis* and *M. tuberculosis*, the orthologous systems have been shown to share certain characteristics and to secrete both sets of substrates [25, 28, 29]. This suggests that the ESX secretion systems have retained conserved mechanisms, and that virulence-associated functions may have evolved subsequently, or be associated with specific substrates.

ESX gene clusters have also been identified in the genomes of closely related actinomycetes outside of the genus *Mycobacterium*, including *Nocardia*, *Streptomyces* and *Corynebacteria* [2, 6]. Furthermore, genes encoding two components of the ESX secretion system, the WXG (or Esx-like) and FtsK/SpoIIIE proteins, have been found clustered in some gram-positive monoderm genera such as *Bacillus*, *Listeria* and *Saccharomyces* [30]. Indeed, it has been suggested that ESX secretion systems occur outside of the *Mycolata* (species containing a mycomembrane-like outer membrane containing mycolic acids, including *Corynebacteria*, *Rhodococci*, *Nocardia* and *Mycobacteria*) and are therefore not typically involved in transmycomembrane secretion [31]. This, together with the absence of an identifiable component responsible for mycomembrane translocation, or an elucidated Type VII secretory mechanism, has generated some controversy, as

some suggest that these are requirements for the designation of the ESX secretion systems as distinct Type VII secretion machineries [32].

Here we investigated the presence and absence of the ESX gene clusters in the genomes of the sequenced mycobacteria and other representative species from the class *Actinobacteria*. The phylogenetic relationship between these and the identified WXG-FtsK clusters of certain monoderm bacteria was determined in order to define the evolutionary history of the Type VII ESX secretion systems. In addition to the five ESX gene clusters which were previously identified, ESX gene clusters were identified on plasmids within several species of mycobacterium, and shown to precede the genomic ESX duplications. A model is proposed for the plasmid-mediated duplication and evolution of the ESX gene clusters.

Results and discussion

ESX gene clusters were identified from the publicly available genome sequences of 60 actinobacterial species, including 40 mycobacterial species, 11 other species from the order *Corynebacteriales* and 9 species selected from the orders *Pseudonocardiales*, *Glycomycetales*, *Micromonosporales*, *Frankiales*, *Streptosporangiales*, *Catenulisporales*, *Streptomyetales*, *Propionibacteriales* and *Kineosporiales* (Table 1). Each genome contains between one and seven ESX gene clusters. The components and arrangement of each ESX gene cluster were determined and are represented in Additional file 1 with three WXG-FtsK clusters from *Staphylococcus aureus*, *Listeria monocytogenes* and *Bacillus subtilis*, identified in the literature as precursors of the ESX gene cluster [30, 31]. The concatenated protein sequences of each ESX gene cluster were aligned and used to generate a phylogeny of the ESX gene clusters using maximum likelihood (ML) and distance methods (Fig. 1 and Additional file 2) using the WXG-FtsK clusters of *S. aureus*, *L. monocytogenes* and *B. subtilis* as the outgroup. The topology of the trees generated by ML and distance methods was conserved, depicting 5 distinct clades, each incorporating one of the *M. tuberculosis* H37Rv ESX gene cluster regions 1 to 5.

ESX gene clusters were identified on plasmids in several mycobacterial species (pMFLV01 in *M. gilvum*, pMKMS01 and pMKMS02 in *M. sp.* KMS, Plasmid01 in *M. sp.* MCS, pMYCCH.01 and pMYCCH.02 in *M. chubuense*, pMYCSM01, pMYCSM02 and pMYCSM03 in *M. smegmatis* JS623, Plasmid 2 in *M. abscessus* sp. *bolletii* and pMyong1 in *M. yongonense*). Four additional mycobacterial plasmid-encoded ESX gene clusters were previously identified by Ummels et al., (2014) [33]. The sequences of three of these, on pRAW from *M. marinum* E11, pMAH135 from *M. avium* subsp. *hominis suis* T135 and pMK12478 from *M. kansasii* ATCC12478, are

Table 1 WXG-FtsK and ESX gene clusters identified in sequenced mycobacterial and selected actinobacterial species

Species	WXG_FtsK	Genomic ESX						Plasmid ESX				
		4	4ev	3	1	2	5	P1	P2	P3	P2'	P5
<i>Bacillus subtilis</i> subsp. <i>subtilis</i> str. 168	x											
<i>Catenulisporea acidiphila</i> DSM43021		xxxx										
<i>Corynebacterium diphtheriae</i> NCTC 13129		x										
<i>Corynebacterium pseudotuberculosis</i> FRC41		x										
<i>Frankia alni</i> ACN14a		x										
<i>Gordonia bronchialis</i> DSM 43247		x										
<i>Janibacter</i> sp. HTCC2649		x										
<i>Kribbella flavida</i> DSM17836		xx										
<i>Listeria monocytogenes</i> L312	x											
<i>M. abscessus</i>		x		x								
<i>M. abscessus</i> subsp. <i>bolletii</i> 50594			x	x					x			
<i>M. africanum</i> GM041182		x		x	x	x	x					
<i>M. avium</i> 104		x		x		x	x					
<i>M. avium</i> paratuberculosis K-10		x		x		x	x					
<i>M. bovis</i> AF2122/97		x		x	x	x	x					
<i>M. bovis</i> BCG Pasteur 1173P2		x		x	x ^b	x	x					
<i>M. canettii</i> CIPT140010059		x		x	x	x	x					
<i>M. chubuense</i>		x			x						xx	
<i>M. colombiense</i> CECT3035 ^a		x		x		x	x					
<i>M. fortuitum</i> subsp. <i>fortuitum</i> DSM 46621 ^a		x	x	x	x							
<i>M. gilvum</i> PYR-GCK		x		x	x			x				
<i>M. indicus pranii</i> MTCC9506		x		x		x	x					
<i>M. intracellulare</i> ATCC13950		x		x		x	x					
<i>M. kansasii</i> ATCC12478 ^a		x		x	x	x	x					
<i>M. leprae</i> TN				x	x		x					
<i>M. ulcerans</i> subsp. <i>liflandii</i> 128FXT		x		x	x		x					
<i>M. marinum</i> M		x		x	x		x					
<i>M. massiliense</i> CCUG48898 ^a		x		x								
<i>M. microti</i> 19422 ^a		x		x	x ^b	x	x					
<i>M. neoaurum</i> VKM Ac-1815D		x		x	x							
<i>M. orygis</i> 112400015		x		x	x	x	x					
<i>M. parascrofulaceum</i> BAA-614 ^a		x		x		x	x		x		x	x
<i>M. phlei</i> RIVM601170 ^a		x		x	x							
<i>M. rhodesiae</i> NBB3 ^a		x		x	x							
<i>M. smegmatis</i> mc ² 155		x		x	x							
<i>M. sp.</i> JDM601		x		x		x	x					
<i>M. sp.</i> JLS		x		x	x							
<i>M. sp.</i> KMS		x		x	x				x		x	
<i>M. sp.</i> MCS		x		x	x					x		
<i>M. sp.</i> MOTT36Y		x		x		x	x					
<i>M. sp.</i> Spyr 1		x		x	x							
<i>M. thermoresistibile</i> 19527		x		x	x							

Table 1 WXG-FtsK and ESX gene clusters identified in sequenced mycobacterial and selected actinobacterial species (Continued)

<i>M. tuberculosis</i> H37Rv	x	x	x	x	x			
<i>M. tusciae</i> JS617 ^a	x	x	x	x ^c		xx	x	
<i>M. ulcerans</i> Agy99	x	x			x			
<i>M. vaccae</i> ATCC 25954	x	x	x	x				
<i>M. vanbaalenii</i> PYR-1	x	x	x					
<i>M. xenopi</i> RIVM700367 ^a	x	x		x	x			
<i>M. yongonense</i> 05-1390	x	x		x	x			x
<i>Nocardia brasiliensis</i> ATCC700358	x	x		x ^c				
<i>Nocardia cyriacigeorgica</i> GUH-2	x	x		x ^c				
<i>Nocardia farcinica</i> IFM 10152	x	x						
<i>Rhodococcus equi</i> 103S	x							
<i>Rhodococcus erythropolis</i> PR4	x							
<i>Rhodococcus opacus</i> B4	x							
<i>Staphylococcus aureus</i> USA300		x						
<i>Streptomyces coelicolor</i> A3 (2)	x							
<i>Saccharopolyspora erythraea</i> NRRL 2338	xx							
<i>Stackebrandtia nassauensis</i> DSM 44728	xxxx							
<i>Streptosporangium roseum</i> DSM 43021	x							
<i>Segniliparus rotundus</i> DSM 44985		x	x					
<i>Salinispora tropica</i> CNB-440	xx							
<i>Tsukamurella paurometabola</i> DSM 20162		x		x ^c				

^aSequencing projects are incomplete (as of 07/2015)^bRD1 deletion within cluster^cAncestral region

publicly available and were included in the phylogenetic analyses. The plasmid-encoded ESX clusters group phylogenetically with some of the ESX gene clusters identified on contigs from the incomplete genome sequences of *M. tusciae* and *M. parascrofulaceum* and together form a subclade of each genomic ESX duplication subsequent to ESX-4 (Fig. 1). The *M. parascrofulaceum* and *M. tusciae* sequencing projects are incomplete, therefore it was not possible to conclusively determine whether the ESX gene clusters identified in these species are plasmid or chromosomally located. However, based on synteny and the phylogenetic clustering of these *M. tusciae* and *M. parascrofulaceum* ESX with the plasmid-encoded ESX clusters, these ESX are predicted to be encoded on plasmids, or to originate directly from plasmid DNA. Sequence alignments indicate that each contig containing a predicted plasmid-located ESX cluster shares several conserved segments, or locally collinear blocks (LCBs), with the ESX-containing plasmids from the same subclade (Additional file 3). This is particularly apparent for sequences containing the subclade of ESX-3, which consist almost entirely of four LCBs, and the subclade of ESX-5. This supports the definition of these *M. tusciae* and *M. parascrofulaceum* ESX gene clusters as plasmid ESX gene clusters. The ESX gene clusters on

the plasmids and *M. tusciae* and *M. parascrofulaceum* contigs, which form outgroups to ESX-1, -2, -3 and -5, were named ESX-P1, -P2, -P2', -P3 and -P5, where "P" indicates the plasmid localisation of the ESX (Table 2). ESX-P1, ESX-P2, ESX-P3 and ESX-P5 form outgroups to the genomic ESX with the same numbers and ESX-P2' branches off prior to ESX-P2. ESX-P1 to -P5 contain all of the core ESX components, including *espG* and *espI* and ESX-P1 also incorporates *EspH*, while *EccA* is absent from ESX-P2.

ESX-4

Orthologs of the ESX-4 gene cluster were identified in all of the mycolic acid producing species from the genera *Mycobacterium*, *Gordonia*, *Nocardia*, *Rhodococcus* and *Corynebacterium*. ESX-4 gene clusters were also identified in the 9 species from the orders *Pseudonocardiales*, *Glycomycetales*, *Micromonosporales*, *Frankiales*, *Streptosporangiales*, *Catenulisporales*, *Streptomycetales*, *Propionibacteriales* and *Kineosporiales* which do not have mycolic acids in their cell envelope. These organisms each contain between one and four copies of the ESX-4 gene cluster. Although the arrangement and components of this gene cluster are well conserved amongst the mycobacterial species; insertions, deletions and rearrangements are common amongst the

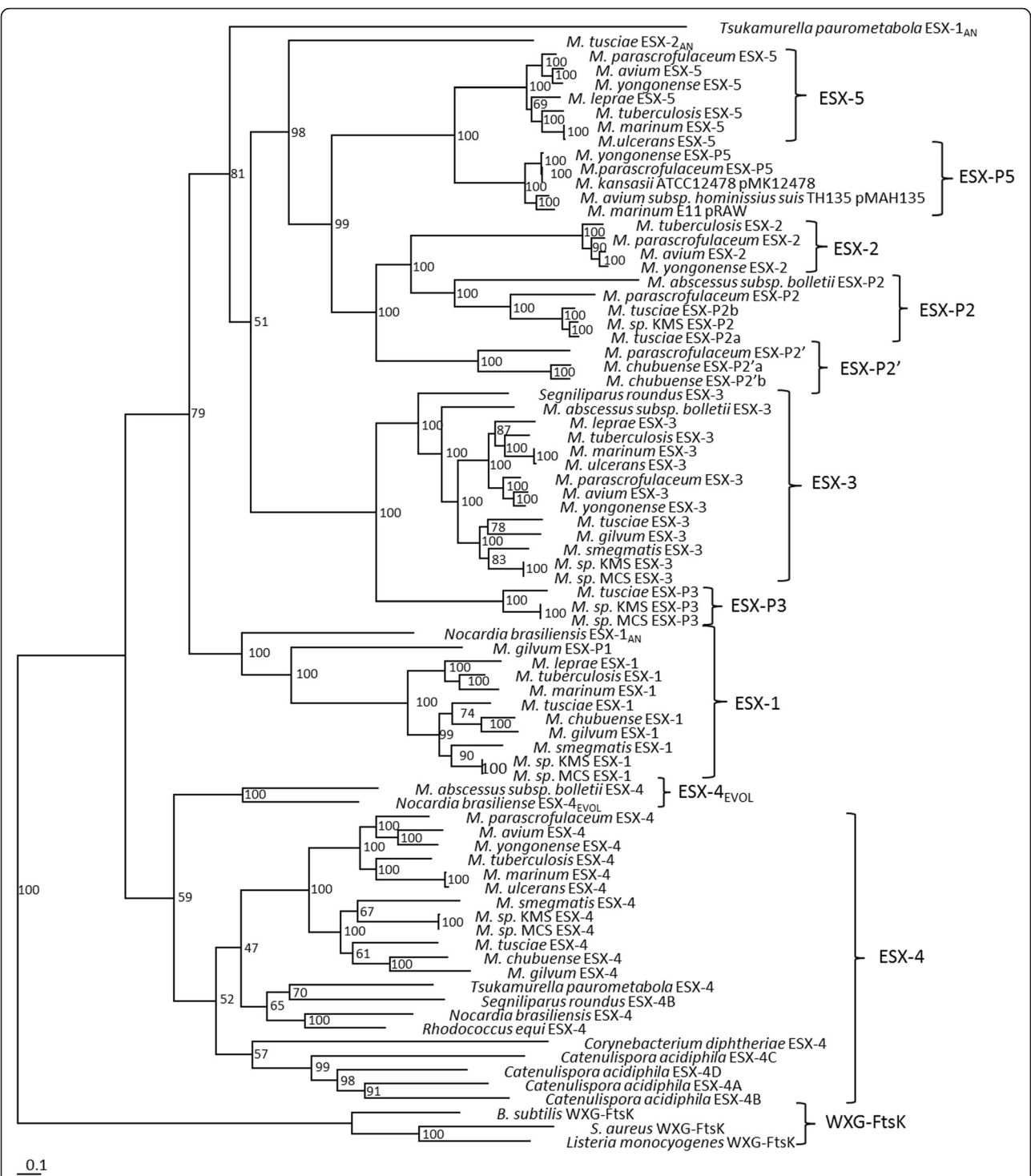


Fig. 1 The phylogeny of the ESX gene cluster. Maximum likelihood phylogeny of representative ESX gene clusters describing the evolution of the ESX gene cluster from its WYG-FtsK cluster progenitor. The ESX gene clusters form five groups, ESX-4, ESX-3, ESX-1, ESX-2 and ESX-5. The plasmid located and ancestral ESX gene clusters form subgroups of each genomic ESX gene cluster. The ESX gene clusters have evolved divergently from a single duplication of ESX-4 to ESX-1 and ESX-3 and then ESX-2 and ESX-5. One hundred subsets were generated for bootstrapping resampling of the data

Table 2 The plasmid-encoded ESX clusters

ESX	Species	Plasmid/contig	Accession number	Size (bp)
P1	<i>M. gilvum</i> PYR-GCK	pMFLV01	NC_009339.1	321,253
P2	<i>M. abscessus</i> subsp. <i>bolletii</i> 50594	plasmid 2	NC_021279.1	97,240
	<i>M. parascrofulaceum</i> BAA-614	contig00115	ADNV01000102.1	21,921
	<i>M. sp</i> KMS	pMKMS01	NC_008703.1	302,089
	<i>M. tusciae</i> JS617	contig 196	NZ_AGJJ01000027.1	108,484
	<i>M. tusciae</i> JS617	contig 209	NZ_AGJJ01000007.1	249,244
P2'	<i>M. chubuense</i>	pMYCCH.01	NC_018022	615,278
	<i>M. chubuense</i>	pMYCCH.02	NC_018023	143,623
	<i>M. parascrofulaceum</i> BAA-614	contig00017	ADNV01000015.1	70,331
P3	<i>M. sp</i> KMS	pMKMS02	NC_008704.1	216,763
	<i>M. sp.</i> MCS	Plasmid1	NC_008147.1	215,075
	<i>M. tusciae</i> JS617	contig 224	NZ_AGJJ01000010.1	218,303
P5	<i>M. avium</i> subsp. <i>hominissuis suis</i> TH135	pMAH135	AP012556	194,711
	<i>M. kansasii</i> ATCC12478	pMK12478	CP006836	144,951
	<i>M. marinum</i> E11	pRAW	HG917973	114,229
	<i>M. parascrofulaceum</i> BAA-614	contig00109	ADNV01000096.1	47,725
	<i>M. yongonense</i> 05-1390	pMyong1	JQ657805	122,976

non-Mycolata. The ESX-4 gene cluster contains genes encoding the FtsK/SpoIIIE protein EccC, and two WXG proteins, EsxU and EsxT, which are present in the FtsK-WXG clusters of *S. aureus*, *L. monocytogenes* and *B. subtilis*. In addition to the WXG-FtsK cluster components, ESX-4 encodes EccD, EccB and MycP, which have been suggested to be involved in a more intricate secretion mechanism to transport proteins into and across the unique and complex outer mycomembrane [34]. However, the presence of the ESX-4 cluster in various non-mycobacterial containing actinobacteria suggests that the secretion system encoded by these gene clusters is not directly involved in mycomembrane translocation. Although the function(s) of ESX-4 have yet to be determined, the presence and maintenance of this gene cluster throughout the mycobacteria and other actinobacteria suggests that it plays an important role in bacterial metabolism. Homologs of the ESX-4 gene cluster components occur in all 5 ESX gene clusters and could represent the proteins required for translocation across the inner membrane. The additional components present in the subsequent ESX duplications may be involved in mycomembrane translocation, be additional substrates, assist in the translocation of additional substrates or facilitate specific mechanisms of those secretion systems.

Phylogenetically associated with the ESX-4 gene cluster is a subgroup of ESX gene clusters which include homologs of the *eccA*, *eccE*, *espG*, *espI*, *pe* and *ppe* genes, in addition to the ESX-4 components. This cluster was identified in the mycolic acid producing species *N. farcinica*, *N. brasiliense*, *N. cyriacigeorgica*, *T. paurometabola*, *S. rotundus*, *M. vaccae*, *M. fortuitum* and *M.*

abscessus subsp. *bolletii*. The arrangement of the genes in this cluster varies between species, but does not resemble any of the *M. tuberculosis* ESX gene clusters. This cluster contains all of the conserved ESX gene cluster components and appears to be an evolutionary intermediate between ESX-4 and the subsequent duplications, and is therefore named ESX-4_{EVOL} (ESX-4 evolved).

ESX-3

ESX-3 present in all of the studied mycobacteria, with the exception of *M. chubuense*, suggesting that ESX-3 is the first ESX duplication in the mycobacterial genome. ESX-3 contains all of the ESX conserved components *eccA* to *E*, *mycP*, *esx* and *pe/ppe* pairs as well as *espG*. Although essential for in vitro growth of *M. tuberculosis*, ESX-3 is not essential in the fast-growing *M. smegmatis* [20]. ESX-3 is involved in iron homeostasis and uptake via the mycobactin pathway [24] and genetic reduction during evolution of the slow-growers may have eliminated the redundancy of ESX-3. Outside of the mycobacteria, ESX-3 was only identified in *S. rotundus* suggesting that ESX-3 was inserted prior to the divergence of *Segniliparus* and *Mycobacterium* from a common ancestor. The presence of three mycobactin genes in the *S. rotundus* ESX-3 furthermore suggests that the association between ESX-3 and iron homeostasis may be conserved. The ancestral mycobacteria *M. abscessus*, *M. abscessus* subsp. *bolletii* and *M. massiliense* contain only ESX-4 (or ESX-4_{evol}) and ESX-3.

ESX-1

ESX-1 is present in all of the other fast-growing mycobacteria; *M. thermoresistibile*, *M. smegmatis* mc²155, *M. neoaurum*, *M. fortuitum*, *M. vanbaalenii*, *M. gilvum*, *M. sp.* Spyr1, *M. vaccae*, *M. rhodesiae*, *M. phlei*, *M. sp.* JLS, *M. sp.* KMS and *M. sp.* MCS; but has been deleted from the genomes of various slow-growing mycobacteria (*M. avium*, *M. avium paratuberculosis*, *M. colombiense*, *M. intracellulare*, *M. parascrofulaceum*, *M. ulcerans*, *M. xenopi*, *M. indicus pranii*, *M. sp.* MOTT36Y and *M. sp.* JDM601), with partial deletions (Region of Deletion 1, RD1) in *M. bovis* BCG and *M. microti*. ESX-1 contains both *espG* and *espI*, and in most cases *eccE* and *mycP* have been inverted along with the insertion of several additional genes. ESX-1 has been implicated in virulence, and its deletion in attenuation of the pathogenic mycobacteria [8, 9]. However, its presence throughout most of the mycobacteria, including non-pathogenic and saprophytic fast-growing organisms, suggests that the primary function of this gene cluster is not virulence, and that the virulence-associated function has evolved more recently in pathogenic organisms.

An additional gene cluster, identified in the non-mycobacterial actinomycetes *N. brasiliense* and *N. cyriacigeorgica* contains all of the components of ESX-4^{EVOL}, but has an operonic arrangement similar to the *M. tuberculosis* ESX gene clusters. This cluster forms a subgroup just outside of the mycobacterial ESX-1 clade and is therefore named ESX-1_{AN} (ancestral ESX-1). An ESX gene cluster with similar arrangements was identified in *T. paurometabola*, but has undergone a transposition event which has resulted in the disruption of *eccC* and deletion of *eccB*. Phylogenetic clustering of this region is not consistent between algorithms and this region is also predicted to be an ESX-1_{AN} cluster, based on synteny.

ESX-2 and ESX-5

ESX-2 and ESX-5 occur only in the slow-growing mycobacteria. ESX-2 contains all of the conserved ESX components including *espG* and *espI* in an operonic structure, while ESX-5 contains only *espG*, but has multiple copies of *pe* and *ppe*, and the insertion of a ferredoxin and a *cyp143* gene. The function(s) of ESX-2 have not been elucidated, and although its duplication correlates evolutionarily with both the slow-growing and pathogenic phenotypes, it has been lost from some of these species (*M. leprae*, *M. marinum*, *M. ulcerans* subsp. *liflandii* and *M. ulcerans*). ESX-5 is the only ESX gene cluster present in all of the slow-growers but absent in all of the fast-growers, and may be the ESX gene cluster most involved in pathogenicity and the slow-growing phenotype [35]. Deletion of this region, however, does not directly increase the growth rate of *M. marinum* or *M. tuberculosis* [18, 36]. ESX-5 has been

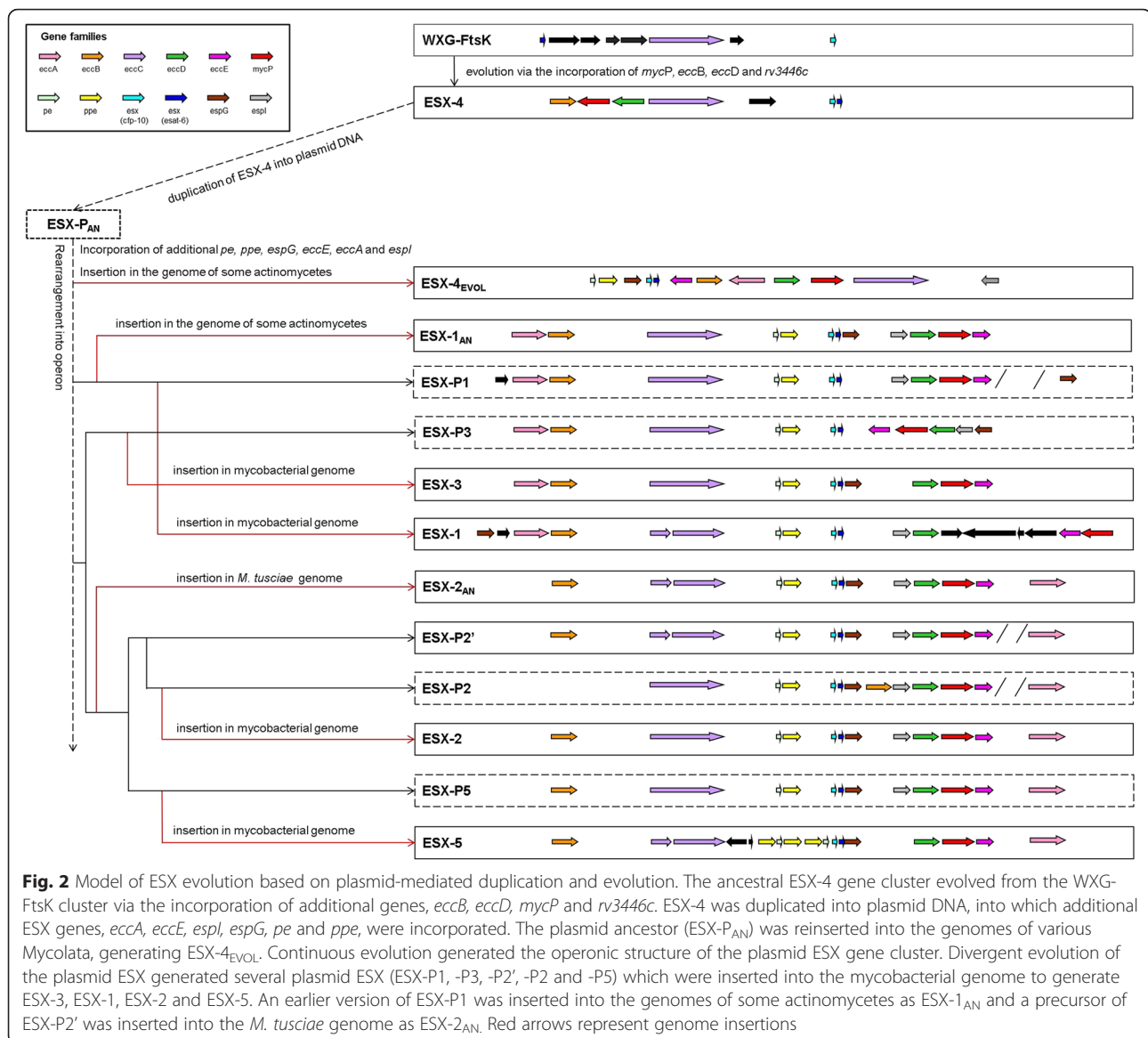
implicated in immune evasion and in the secretion of the PE and PPE proteins [36, 37]. Only ESX-5 contains multiple copies of the *pe* and *ppe* genes, the numbers of which vary between species, and its evolution is predicted to have preceded the expansion of these gene families [37].

M. tusciae contains an ESX cluster, ESX-2_{AN} (ancestral ESX-2), which contains all of the ESX-2 components and precedes both the ESX-2 and ESX-5 clades, as well as ESX-P2; -P2 and -P5 gene clusters. *M. tusciae* is a slow-growing mycobacterium which, based on 16S rDNA sequencing, clusters with the fast-growing mycobacteria and is most closely related to the fast-growing mycobacteria *M. farcinogenes*, *M. komossense* and *M. aichiense* [38]. The correlation between the presence of an ESX-2/5-like cluster and a slow growth-rate might imply that *M. tusciae* is an evolutionary intermediate between the fast- and slow-growing mycobacteria. The mycolic acid composition of the cell membrane of *M. tusciae* most closely resembles that of the *M. avium* complex and *M. parascrofulaceum* [38] suggesting that the different ESX secretion systems may have evolved with changes in the mycomembrane structure; as reflected in the role of ESX-5 in maintaining selective mycomembrane permeability in the slow growing pathogenic *M. tuberculosis* and *M. marinum* species [19]. Investigation of the potential association between these two ESX clusters, mycomembrane structure and growth rate may provide important information regarding the evolution of the often pathogenic, slow-growing mycobacteria.

Plasmid-mediated ESX evolution

The duplication and evolution of the ESX gene clusters and their secretion systems have clearly impacted on the evolution, diversity and success of the mycobacteria. The identification of ESX gene clusters on several plasmids within the mycobacteria, and their phylogenetic association with each of the genomic ESX gene clusters, provides novel insight into the mechanism of ESX evolution suggesting that the duplication and diversification of these clusters was plasmid-mediated. The presence of multiple plasmid copies within a single organism facilitates diversification by allowing the coevolution of various ESX clusters simultaneously. The plasmid localisation furthermore facilitates the loss of deleterious effects, while the incorporation of beneficial plasmid DNA into the genome allows permanent retention and might be selected for. We propose a model for the plasmid-mediated duplication and evolution of the ESX gene clusters (Fig. 2).

Based on this model, the FtsK-WXG cluster present in the *Firmicutes* evolved to form the ESX-4 gene cluster, through the incorporation of *eccB*, *eccD* and *mycP*, during the evolution of the actinobacteria; resulting in the presence of ESX-4 in the genomes of various actinobacterial



species. A copy of ESX-4 has been incorporated into plasmid DNA after the divergence of the genera *Corynebacterium* and *Rhodococcus*. The additional ESX components, *eccA*, *eccE*, *espG*, *pe* and *ppe*, were incorporated into this plasmid-located cluster (ESX-P_{AN}), which was subsequently incorporated into the genomes of some species, including *Nocardia* spp., *T. paurometabola*, *S. rotundus* and *M. abscessus* subsp. *bolletii*, as ESX-4_{EVOL}. The variation in the arrangement and sequences of the genes in these clusters may represent independent insertions at different evolutionary time points. The presence of both ESX-4 and ESX-4_{EVOL} in some species implies that ESX-4_{EVOL} is a duplication of the ESX-4 cluster, and has not evolved directly from it. ESX-1, -2, -3, and 5 have evolved from a single duplication of ESX-4. The presence of all of the conserved ESX components in ESX-4_{EVOL} suggests

that it evolved from the same progenitor and that ESX-4_{EVOL} is an intermediate between ESX-4 and ESX-1, -2, -3 and -5. Continual evolution of this plasmid ESX gene cluster generated the operonic structure characteristic of the mycobacterial ESX gene cluster duplications. Plasmid precursors of the four duplications, ESX-P1, -P3, -P2', -P2 and -P5, have evolved simultaneously by divergence of the common plasmid ancestor, after which genome insertions generated the genomic ESX-1, -2, -3 and -5 clusters.

It appears furthermore, that these plasmids may be able to transfer between mycobacterial species. The pRAW, pMyong1, pMK12478 and pMAH135 plasmids, which contain ESX-P5, were also shown to contain components of a Type-IV secretions system and a *traA*/relaxase gene; which are required for conjugation of the plasmid between some slow-growing mycobacterial species [33].

ESX-associated evolution of the mycobacteria

A phylogenetic analysis of the mycobacteria and related actinomycetes based on their ESX gene clusters was done

using the concatenated protein sequences of all of the ESX gene clusters of each species (Fig. 3). The *Mycolata* have evolved from a single gram-positive monoderm

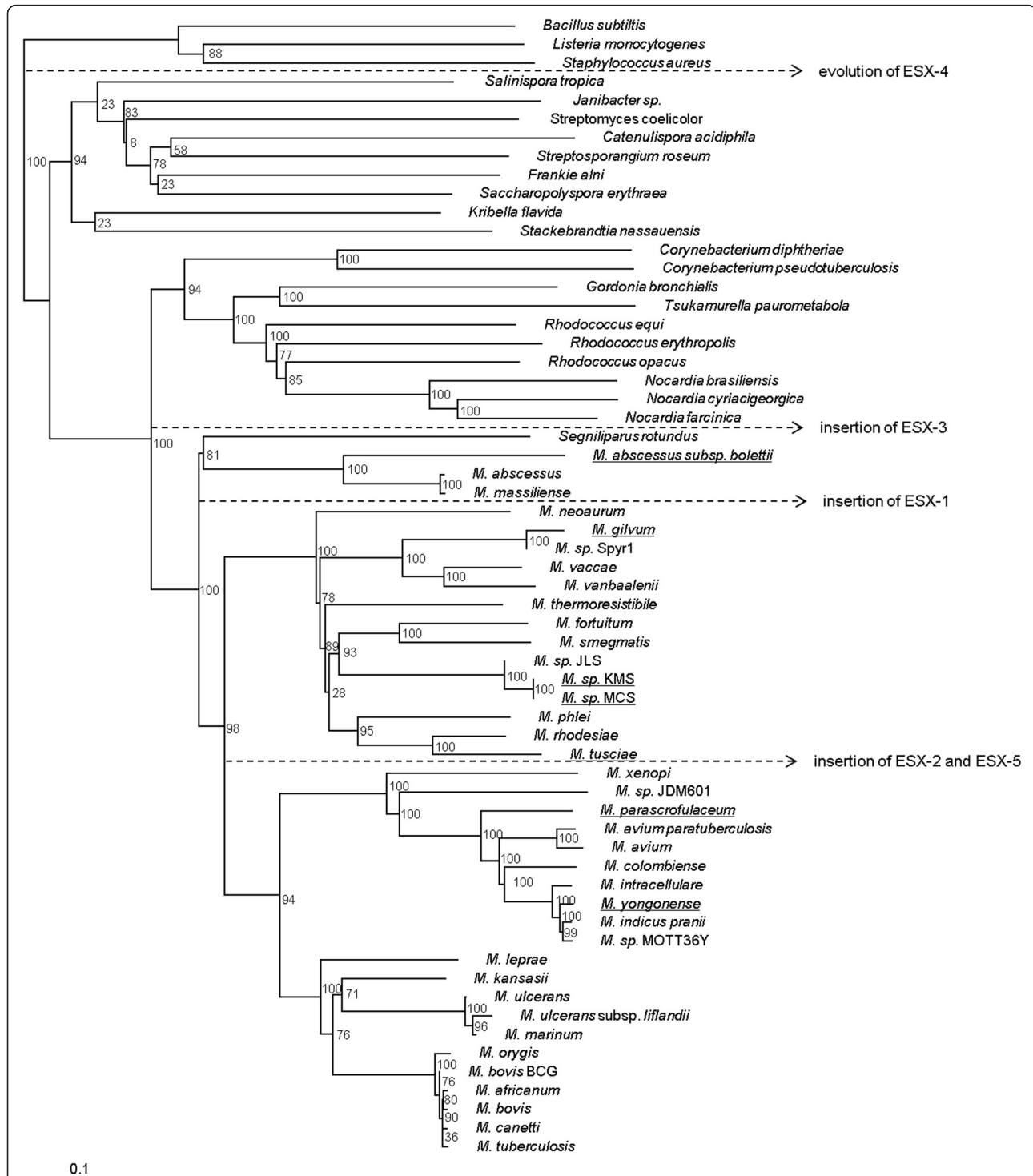


Fig. 3 The phylogeny of the mycobacteria based on ESX duplication and evolution. Maximum likelihood phylogeny describing the evolution of the mycobacteria based on the concatenated ESX gene cluster amino acid sequences from each species. ESX duplication and deletion events influenced the evolution and diversification of the mycobacteria as described in the text. Species which contain plasmid ESX gene clusters are underlined. One thousand subsets were generated for bootstrapping resampling of the data

ancestor into two groups, those which contain only ESX-4, ESX-4_{EVOL} and ESX-1_{AN}, the non-mycobacterial actinomycetes; and those which also contain an ESX-3 gene cluster, which with the exception of *S. rotundus*, consist of the mycobacteria. *S. rotundus* contains ESX-4_{EVOL} and ESX-3, while all of the mycobacteria contain at least ESX-4 and ESX-3, with the exception of *M. leprae* which has lost ESX-4. ESX-1 was incorporated in the mycobacterial genome after the divergence of *M. abscessus* and *M. massiliense*, and is present in all of the other fast-growing mycobacteria. However, an ESX-1-like cluster (ESX-1_{AN}) is also present in some *Nocardia* spp.. ESX-1_{AN} predates ESX-P1 and was likely incorporated into the genome from an earlier form of ESX-P1, after the divergence of the mycobacteria. The presence of ESX-1_{AN} in the absence of ESX-3 in some *Nocardia* species, and the presence of ESX-3 in the absence of ESX-1 in *M. abscessus* ssp., *M. massiliense* and *S. rotundus* suggests that these plasmid clusters evolved simultaneously in an ancestral species, and were inserted into the genomes of the different organisms at different times. The role of ESX-1 in conjugal DNA transfer in *M. smegmatis* [25, 26] may be linked to its origin in plasmid DNA, where it may have facilitated the transfer of the plasmid during cell division.

The presence of ESX-2 and ESX-5 marks the emergence of the slow-growing mycobacteria. ESX-2 and ESX-5 evolved from a common ancestral plasmid-ESX, which diverged to produce ESX-P2' and ESX-P2; and -P5. ESX-2 and ESX-5 were integrated into the mycobacterial genome with the divergence of the slow-growing mycobacteria, however the presence of ESX-2_{AN}, ESX-P2' and ESX-P2 in various fast-growing mycobacteria attests to the presence of these precursors earlier in mycobacterial evolution. The *M. avium* complex can be distinguished by the transposition of EccB₂ and EccC₂. ESX-2 was deleted from a precursor of *M. ulcerans*, *M. ulcerans* subsp. *liflandii* and *M. marinum*. ESX-1 was deleted from the genomes of slow-growing mycobacteria on numerous occasions. *M. kansasii* and the *M. tuberculosis* complex have retained all five ESX gene clusters, with the exception of *M. bovis* BCG and *M. microti* (not shown), which contain the previously described RD1 deletions in ESX-1 [3, 4, 39–41]. *M. leprae*, which has undergone extensive gene reduction, has retained only ESX-3, -1 and -5 and *M. ulcerans* has retained only ESX-4, -3 and -5.

Conclusion

The distinctive cell envelope of mycobacteria, characterised by the highly impermeable outer mycomembrane peptidoglycan-arabinogalactan-mycolic acid matrix [6], provides a protective barrier against extracellular stresses, but also presents an obstacle to the export of proteins and acquisition of nutrients. Although mycobacteria possess

both Sec and Tat secretion systems, which translocate proteins across the inner membrane, the ESX, or Type VII, secretion systems are the first mechanism proposed for the secretion of proteins into and across the mycomembrane. This study explored the evolution of the mycobacterial Type VII ESX gene clusters from the WXG-FtsK cluster in *S. aureus*, *L. monocytogenes* and *B. subtilis* to the 5 ESX gene clusters in *M. tuberculosis*. The ancestral ESX gene cluster (ESX-4) was identified in several non-mycobacterial producing actinobacteria as well as the non-mycobacterial *Corynebacteriales*. Between two and seven ESX gene clusters were identified in each mycobacterial species. A novel ESX gene cluster, ESX-4_{EVOL}, was identified in some non-mycobacterial myco-membrane containing actinomycetes and *M. abscessus* subsp. *bolletii*. ESX-4_{EVOL} contains all of the conserved components of the ESX and appears to be a precursor of the mycobacterial ESX duplications. Plasmid-encoded precursor ESX were identified for each of the genomic ESX-3, -1, -2 and -5 gene clusters and a novel plasmid-mediated mechanism of ESX duplication and evolution proposed. The presence and absence of the ESX gene clusters in the mycobacteria redefines the order of duplication of the ESX gene clusters in the mycobacteria as ESX-4, ESX-3, ESX-1 and then ESX-2 and ESX-5. The influence of the various ESX gene clusters on vital biological and virulence-related functions has clearly influenced the diversification and success of the various mycobacteria, and their evolution from the non-pathogenic fast-growing saprophytic to the slow-growing pathogenic organisms.

Methods

Genome sequence data

All protein and DNA sequence information was obtained from publicly available finished and unfinished genome sequencing information. The genomes of 40 mycobacterial species, 11 other species from the order *Corynebacteriales*, nine species selected from the orders *Pseudonocardiales*, *Glycomycetales*, *Frankiales*, *Micromonosporales*, *Streptosporangiales*, *Catenulisporales*, *Streptomycetales*, *Propionibacteriales* and *Kineosporiales* and 3 gram-positive monoderm species containing WXG-FtsK clusters (Table 1), were analysed.

Comparative genomic analyses

The *M. tuberculosis* H37Rv ESX protein sequences of interest were used as templates to identify orthologous ESX protein and gene sequences. Blast similarity searches, blastn, tblastn and blastp [42], were done using NCBI Blast and the genome sequence databases listed in Additional file 4. Adjacent genomic regions were searched for additional ESX genes to determine clustering and arrangement of genes; for unfinished genomes

in contig format this was not always possible and gene cluster arrangement was assumed based on sequence identity and anticipated arrangement. Large intergenic regions were searched for gene insertions using blastx analyses [43].

Phylogenetic analyses

Annotated protein sequences were obtained from the protein sequence databases. The protein sequences of conserved components of each ESX gene cluster (EccA, EccB, EccC, EccD, EccE, PE(s), PPE(s), Esx (CFP-10-like), Esx (ESAT-6-like), EspG, EspI, MycP, Rv3446c, EspH, EspJ, EspK, EspL, EspB, Cyp143 and Ferredoxin) were concatenated. Multiple sequence alignments of all concatenated ESX gene cluster protein sequences were done with Clustal W 2.0 [44, 45] using the Bioedit Sequence Alignment Editor version 7.1.3.0 [46]. Similarly, multiple sequence alignments of a single sequence composed of all of the combined ESX gene cluster protein sequences, from each species, were done. Phylogenetic trees were determined by distance and maximum likelihood analyses using SeaView Version 4.4.2 [47]. Distance analysis was done using the observed neighbour-joining method with 10000 bootstrap replicates. Maximum likelihood phylogenies were generated using PhyML [48] with the JTT (Jones Taylor Thornton) algorithm [49], using model-given amino acid equilibrium frequencies, specifying no invariable sites and no across site variation. Nearest-neighbor interchange tree searching operations were used with a BioNJ starting tree. The WXXG-FtsK cluster sequences from *S. aureus*, *L. monocytogenes* and *B. subtilis* were defined as the outgroup. The *M. microti* ESX clusters were omitted from the phylogenetic analyses as protein annotations were not available.

Plasmid and contig sequence alignments

Plasmid and contig sequences were obtained from the NCBI (Additional file 4) and alignments of the plasmid and contig sequences containing each subgroup of ESX gene cluster were done using the progressiveMauve algorithm of the Mauve 2.3.1 Genome Alignment Visualisation software [50].

Availability of supporting data

All supporting data are included as Additional files 1, 2, 3 and 4.

Additional files

Additional file 1: The ESX gene clusters of sequenced mycobacteria and selected actinobacterial species. (DOCX 3463 kb)

Additional file 2: Neighbour joining phylogeny of the ESX gene clusters using 10000 bootstrap replicates. (PNG 33 kb)

Additional file 3: Mauve sequence alignments of ESX-P containing plasmids and contigs which are predicted to be plasmids.

Additional file 4: Analysed genomes. (XLSX 20 kb)

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

MNF participated in the design of the study, carried out the analyses and wrote the initial draft of the manuscript. RMW, SLS and PDvH participated in the design of the study and advised on computational methodologies and the interpretation of data. NGvP conceived of and designed the study and advised on computational methodologies and the interpretation of data. All authors contributed to, read and approved the final manuscript.

Received: 30 November 2015 Accepted: 6 March 2016

Published online: 15 March 2016

References

1. Tekala F, Gordon SV, Garnier T, Brosch R, Barrell BG, Cole ST. Analysis of the proteome of *Mycobacterium tuberculosis* in silico. *Tuber Lung Dis*. 1999;79:329–42.
2. Gey Van Pittius NC, Gamielien J, Hide W, Brown GD, Siezen RJ, Beyers AD. The ESAT-6 gene cluster of *Mycobacterium tuberculosis* and other high G+C Gram-positive bacteria. *Genome Biol*. 2001;2:RESEARCH0044.
3. Mahairas GG, Sabo PJ, Hickey MJ, Singh DC, Stover CK. Molecular analysis of genetic differences between *Mycobacterium bovis* BCG and virulent *M. bovis*. *J Bacteriol*. 1996;178:1274–82.
4. Behr MA, Wilson MA, Gill WP, Salamon H, Schoolnik GK, Rane S, Small PM. Comparative genomics of BCG vaccines by whole-genome DNA microarray. *Science*. 1999;284:1520–3.
5. Brosch R, Gordon SV, Buchrieser C, Pym AS, Garnier T, Cole ST. Comparative genomics uncovers large tandem chromosomal duplications in *Mycobacterium bovis* BCG Pasteur. *Yeast*. 2000;17:111–23.
6. Abdallah AM, Gey van Pittius NC, Champion PA, Cox J, Luirink J, Vandenbroucke-Grauls CM, Appelmeijer BJ, Bitter W. Type VII secretion—mycobacteria show the way. *Nat Rev*. 2007;5:883–91.
7. Bitter W, Houben EN, Bottai D, Brodin P, Brown EJ, Cox JS, Derbyshire K, Fortune SM, Gao LY, Liu J, Gey van Pittius NC, Pym AS, Rubin EJ, Sherman DR, Cole ST, Brosch R. Systematic genetic nomenclature for type VII secretion systems. *PLoS Pathog*. 2009;5:e1000507.
8. Hsu T, Hingley-Wilson SM, Chen B, Chen M, Dai AZ, Morin PM, Marks CB, Padiyar J, Goulding C, Gingery M, Eisenberg D, Russell RG, Derrick SC, Collins FM, Morris SL, King CH, Jacobs Jr WR. The primary mechanism of attenuation of bacillus Calmette-Guérin is a loss of secreted lytic function required for invasion of lung interstitial tissue. *Proc Natl Acad Sci U S A*. 2003;100:12420–5.
9. Lewis KN, Liao R, Guinn KM, Hickey MJ, Smith S, Behr MA, Sherman DR. Deletion of RD1 from *Mycobacterium tuberculosis* mimics bacille Calmette-Guérin attenuation. *J Infect Dis*. 2003;187:117–23.
10. Pym AS, Brodin P, Majlessi L, Brosch R, Demangel C, Williams A, Griffiths KE, Marchal G, Leclerc C, Cole ST. Recombinant BCG exporting ESAT-6 confers enhanced protection against tuberculosis. *Nat Med*. 2003;9:533–9.
11. MacGurn JA, Cox JS. A genetic screen for *Mycobacterium tuberculosis* mutants defective for phagosome maturation arrest identifies components of the ESX-1 secretion system. *Infect Immun*. 2007;75:2668–78.
12. Samten B, Wang X, Barnes PF. *Mycobacterium tuberculosis* ESX-1 system-secreted protein ESAT-6 but not CFP10 inhibits human T-cell immune responses. *Tuberculosis (Edinb)*. 2009;89 Suppl 1:S74–6.
13. de Jonge MI, Pehau-Arnaudet G, Fretz MM, Romain F, Bottai D, Brodin P, Honore N, Marchal G, Jiskoot W, England P, Cole ST, Brosch R. ESAT-6 from *Mycobacterium tuberculosis* dissociates from its putative chaperone CFP-10 under acidic conditions and exhibits membrane-lysing activity. *J Bacteriol*. 2007;189:6028–34.
14. Smith J, Manoranjan J, Pan M, Bohsali A, Xu J, Liu J, McDonald KL, Szyk A, LaRonde-LeBlanc N, Gao LY. Evidence for pore formation in host cell membranes by ESX-1-secreted ESAT-6 and its role in *Mycobacterium marinum* escape from the vacuole. *Infect Immun*. 2008;76:5478–87.
15. Abdallah AM, Bestebroer J, Savage ND, de Punder K, van Zon M, Wilson L, Korbbe CJ, van der Sar AM, Ottenhoff TH, van der Wel NN, Bitter W, Peters PJ.

- Mycobacterial secretion systems ESX-1 and ESX-5 play distinct roles in host cell death and inflammasome activation. *J Immunol.* 2011;187:4744–53.
16. Houben D, Demangel C, van Ingen J, Perez J, Baldeon L, Abdallah AM, Caleechurn L, Bottai D, van Zon M, de Punder K, van der Laan T, Kant A, Bossers-de Vries R, Willemssen P, Bitter W, van Soolingen D, Brosch R, van der Wel N, Peters PJ. ESX-1-mediated translocation to the cytosol controls virulence of mycobacteria. *Cell Microbiol.* 2012;4:1287–98.
 17. Abdallah AM, Verboom T, Hannes F, Safi M, Strong M, Eisenberg D, Musters RJ, Vandenbroucke-Grauls CM, Appelmek BJ, Luirink J, Bitter W. A specific secretion system mediates PPE41 transport in pathogenic mycobacteria. *Mol Microbiol.* 2006;62:667–79.
 18. Weerdenburg EM, Abdallah AM, Mitra S, de Punder K, van der Wel NN, Bird S, Appelmek BJ, Bitter W, van der Sar AM. ESX-5-deficient *Mycobacterium marinum* is hypervirulent in adult zebrafish. *Cell Microbiol.* 2012;14:728–39.
 19. Ates LS, Ummels R, Commandeur S, van der Weerd R, Sparrius M, Weerdenburg E, Alber M, Kalscheuer R, Piersma SR, Abdallah AM, Abd El Ghany M, Abdel-Haleem AM, Pain A, Jiménez CR, Bitter W, Houben ENG. Essential Role of the ESX-5 Secretion System in Outer Membrane Permeability of Pathogenic Mycobacteria. *PLoS Genet.* 2015;11:e1005190.
 20. Sassetti CM, Boyd DH, Rubin EJ. Genes required for mycobacterial growth defined by high density mutagenesis. *Mol Microbiol.* 2003;48:77–84.
 21. Serafini A, Boldrin F, Palu G, Manganello R. Characterization of a *Mycobacterium tuberculosis* ESX-3 conditional mutant: essentiality and rescue by iron and zinc. *J Bacteriol.* 2009;191:6340–4.
 22. Maciag A, Dainese E, Rodriguez GM, Milano A, Prowedi R, Pasca MR, Smith I, Palu G, Riccardi G, Manganello R. Global analysis of the *Mycobacterium tuberculosis* Zur (FurB) regulon. *J Bacteriol.* 2007;189:730–40.
 23. Rodriguez GM, Voskuil MI, Gold B, Schoolnik GK, Smith I. IdeR, An essential gene in *Mycobacterium tuberculosis*: role of IdeR in iron-dependent gene expression, iron metabolism, and oxidative stress response. *Infect Immun.* 2002;70:3371–81.
 24. Siegrist MS, Unnikrishnan M, McConnell MJ, Borowsky M, Cheng TY, Siddiqi N, Fortune SM, Moody DB, Rubin EJ. Mycobacterial Esx-3 is required for mycobactin-mediated iron acquisition. *Proc Natl Acad Sci U S A.* 2009;106:18792–7.
 25. Flint JL, Kowalski JC, Karnati PK, Derbyshire KM. The RD1 virulence locus of *Mycobacterium tuberculosis* regulates DNA transfer in *Mycobacterium smegmatis*. *Proc Natl Acad Sci U S A.* 2004;101:12598–603.
 26. Coros A, Callahan B, Battaglioli E, Derbyshire KM. The specialized secretory apparatus ESX-1 is essential for DNA transfer in *Mycobacterium smegmatis*. *Mol Microbiol.* 2008;69:794–808.
 27. Maciag A, Piazza A, Riccardi G, Milano A. Transcriptional analysis of ESAT-6 cluster 3 in *Mycobacterium smegmatis*. *BMC Microbiol.* 2009;9:48.
 28. Wirth SE, Krywy JA, Aldridge BB, Fortune SM, Fernandez-Suarez M, Gray TA, Derbyshire KM. Polar assembly and scaffolding proteins of the virulence-associated ESX-1 secretory apparatus in mycobacteria. *Mol Microbiol.* 2012;83:654–64.
 29. Converse SE, Cox JS. A protein secretion pathway critical for *Mycobacterium tuberculosis* virulence is conserved and functional in *Mycobacterium smegmatis*. *J Bacteriol.* 2005;187:1238–45.
 30. Pallen MJ. The ESAT-6/WXG100 superfamily – and a new Gram-positive secretion system? *Trends Microbiol.* 2002;10:209–12.
 31. Sutcliffe IC. New insights into the distribution of WXG100 protein secretion systems. *Antonie Van Leeuwenhoek.* 2011;99:127–31.
 32. Desvaux M, Hebraud M, Talon R, Henderson IR. Outer membrane translocation: numerical protein secretion nomenclature in question in mycobacteria. *Trends Microbiol.* 2009;17:338–40.
 33. Ummels R, Abdallah AM, Kuiper V, Ajjoud A, Sparrius M, Naeem R, Spaink HP, van Soolingen D, Pain A, Bitter W. Identification of a novel conjugative plasmid in mycobacteria that requires both type IV and type VII secretion. *MBio.* 2014;5:e01744–14.
 34. Brodin P, Majlessi L, Marsollier L, de Jonge MI, Bottai D, Demangel C, Hinds J, Neyrolles O, Butcher PD, Leclerc C, Cole ST, Brosch R. Dissection of ESAT-6 system 1 of *Mycobacterium tuberculosis* and impact on immunogenicity and virulence. *Infect Immun.* 2006;74:88–98.
 35. Abdallah AM, Savage ND, van Zon M, Wilson L, Vandenbroucke-Grauls CM, van der Wel NN, Ottenhoff TH, Bitter W. The ESX-5 secretion system of *Mycobacterium marinum* modulates the macrophage response. *J Immunol.* 2008;181:7166–75.
 36. Bottai D, Di Luca M, Majlessi L, Frigui W, Simeone R, Sayes F, Bitter W, Brennan MJ, Leclerc C, Batoni G, Campa M, Brosch R, Esin S. Disruption of the ESX-5 system of *Mycobacterium tuberculosis* causes loss of PPE protein secretion, reduction of cell wall integrity and strong attenuation. *Mol Microbiol.* 2012;83:1195–209.
 37. van Pittius NC, Sampson SL, Lee H, Kim Y, van Helden PD, Warren RM. Evolution and expansion of the *Mycobacterium tuberculosis* PE and PPE multigene families and their association with the duplication of the ESAT-6 (esx) gene cluster regions. *BMC Evol Biol.* 2006;6:95.
 38. Tortoli E, Kroppenstedt RM, Bartoloni A, Caroli G, Jan I, Pawlowski J, Emmer S. *Mycobacterium tusciae* sp. nov. *Int J Syst Bacteriol.* 1999;49 Pt 4:1839–44.
 39. van Soolingen D, van der Zanden AG, de Haas PE, Noordhoek GT, Kiers A, Foudraïne NA, Portaels F, Kolk AH, Kremer K, van Embden JD. Diagnosis of *Mycobacterium microti* infections among humans by using novel genetic markers. *J Clin Microbiol.* 1998;36:1840–5.
 40. Brodin P, Eiglmeier K, Marmiesse M, Billault A, Garnier T, Niemann S, Cole ST, Brosch R. Bacterial artificial chromosome-based comparative genomic analysis identifies *Mycobacterium microti* as a natural ESAT-6 deletion mutant. *Infect Immun.* 2002;70:5568–78.
 41. Gordon SV, Brosch R, Billault A, Garnier T, Eiglmeier K, Cole ST. Identification of variable regions in the genomes of tubercle bacilli using bacterial artificial chromosome arrays. *Mol Microbiol.* 1999;32:643–55.
 42. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990;215:403–10.
 43. Gish W, States DJ. Identification of protein coding regions by database similarity search. *Nat Genet.* 1993;3:266–72.
 44. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 1994;22:4673–80.
 45. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. Clustal W and Clustal X version 2.0. *Bioinformatics.* 2007;23:2947–8.
 46. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser.* 1999;41:95–8.
 47. Gouy M, Guindon S, Gascuel O. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol.* 2010;27:221–4.
 48. Guindon S, Delsuc F, Dufayard JF, Gascuel O. Estimating Maximum Likelihood Phylogenies with PhyML. *Volume 537.* 2009.
 49. Jones DT, Taylor WR, Thornton JM. The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci.* 1992;8:275–82.
 50. Darling ACE, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 2004;14:1394–403.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

